1. INTRODUCTION

1.1 The case-control study in cancer epidemiology

1.2 Objectives

1.3 Strengths

1.4 Limitations

1.5 Planning

1.6 Implementation

1.7 Interpretation

CHAPTER I

INTRODUCTION[1]

Only two decades ago the case-control study was an oddity; it was rarely performed, poorly understood and, perhaps for these reasons, not highly regarded. But this type of study design has increased steadily in use and in stature and today it is an important and perhaps the dominant form of analytical research in epidemiology, especially in cancer epidemiology. Nonetheless, as a form of research the case-control study continues to offer a paradox: compared with other analytical designs it is a rapid and efficient way to evaluate a hypothesis. On the other hand, despite its practicality, the case-control study is not simplistic and it cannot be done well without considerable planning. Indeed, a case-control study is perhaps the most challenging to design and conduct in such a way that bias is avoided. Our limited understanding of this difficult study design and its many subtleties should serve as a warning – these studies must be designed and analysed carefully with a thorough appreciation of their difficulties. This warning should also be heeded by the many critics of the case-control design. General criticisms of the design itself too often reflect a lack of appreciation of the same complexities which make these studies difficult to perform properly.

The two major areas where a case-control study presents difficulties are in the selection of a control group, and in dealing with confounding and interaction as part of the analysis. This monograph deals mainly with the analysis of case-control studies and with related quantitative issues. This introductory chapter has different objectives: (1) to give a perspective on the place of the case-control study in cancer epidemiology; (2) to describe the major strengths and limitations of the approach; (3) to describe some aspects of the planning and conduct of a case-control study; and (4) to discuss matching, a major issue in designating the control group.

1.1 The case-control study in cancer epidemiology

Definition

A case-control study (case-referent study, case-compeer study or retrospective study) is an investigation into the extent to which persons selected because they have

[1] Prepared by Philip Cole, M.D., Dr. P.H., Division of Epidemiology, Department of Public Health and the Comprehensive Cancer Center of the University of Alabama in Birmingham, USA. Supported by a contract (N01-CO-55195) from the International Cancer Research Data Bank, National Cancer Institute, USA, to the International Agency for Research on Cancer, Lyon, France.

a specific disease (the cases) and comparable persons who do not have the disease (the controls) have been exposed to the disease's possible risk factors in order to evaluate the hypothesis that one or more of these is a cause of the disease. This definition requires considerable expansion to provide a picture of all the major aspects of such studies and the common variations; some of the more important deserve mention.

First, while case-control studies usually include only one case group and one control group, there are three common departures from this situation. For efficiency an investigator may decide to study simultaneously, and in the same way, two or more diseases whose risk factors are thought to be similar. For example, we recently simultaneously studied cancer of the endometrium (Elwood et al., 1977) and benign breast disease (Cole, Elwood & Kaplan, 1978). In addition to the operational efficiencies of such "multi-disease" studies, the control groups may be able to be combined to give each case-control comparison increased power. (Of course, a multi-disease study could be considered as a series of case-control studies, each consisting of two groups.)

There are two more ways in which the use of more than one case series could be useful. In one, cases of a second cancer known to be caused by one of the factors under study could be included. If the factor is also found to be related to the second cancer, that case group would have served as a "positive control", revealing the sensitivity of the study. In another, several case series are included, but only one cancer is found to be related to the factor under study, and thus the other cancers would have a "negative control" function; this would provide some evidence that the association with the cancer of primary interest was not merely reflecting a built-in aspect of the study design.

The second way in which the number of groups is increased beyond two is rarely used in the study of cancer; but for some diseases, such as arteriosclerosis, hypertension and some mental illness, it may be useful to deal with a group of "para-cases" i.e., subjects who are intermediate between the clearly ill and the clearly healthy. The decision to designate such an intermediate group might, in fact, be made when the study is analysed.

The third, and most common, way in which case-control studies are expanded is by the use of more than one control group. Indeed, it has been suggested that a case-control study requires at least two control groups to minimize the possibility of accepting a false result (Ibrahim & Spitzer, 1979); the rationale is that if the same result is not achieved in the two case-control comparisons, both the apparent results are suspect. Inclusion of a second control group may, however, increase the cost and duration of a study by about 50% and this may not be worthwhile. Furthermore, it may be difficult to judge whether or not the results of the two comparisons are mutually supportive.

Usually, it seems judicious to use a single control group – the one which seems best suited to the needs of the particular study. But, there are two common circumstances in which a second control group may be indicated: (1) in the study of a cancer about which so little is known that no strong preference for one type of control group can be defended; (2) in the situation where one desirable control group has a specific deficiency which can be overcome by another desirable group. For example, in a case-control study to evaluate the hypothesis that tonsillectomy causes Hodgkin's disease, Gutensohn et al. (1975) wanted to control potential confounding by socioeconomic class in the study design. This presented a problem since it was not clear whether it was necessary to control for socioeconomic class in childhood or in adulthood or in both. They

therefore used two control groups – the siblings and the spouses of the cases. It is useful to remember that if the two different case-control comparisons give different results the study is not necessarily uninterpretable. The explanation of the discrepancy, if one can be deduced, may be very informative. For example, in the study just mentioned, the relative risk of Hodgkin's disease among tonsillectomized persons was 1.4 using the sibling controls, but 3.1 using the spouses. This suggests that some factor(s), which is a correlate of the probability of having a tonsillectomy, which differs between spouses and which is over-controlled for by the use of sibling controls, is a cause of Hodgkin's disease. Thus, the hypothesis emerges, though not exclusively from this finding, that an aspect of lifestyle during childhood – perhaps the pattern of exposure to infectious agents – is a cause of Hodgkin's disease (Gutensohn & Cole, 1977).

The modes of analysis presented in this monograph relate exclusively to the comparison of a single group of cases and a single group of controls; simultaneous multi-group comparisons are not addressed. This, however, does not prevent use of the techniques presented for the analysis of a study with, say, two control groups. Each case group-control group comparison can be analysed using these techniques and the results "pooled" in a subjective way at the interpretation stage. Also, if the decision is made to pool the data from the control groups, one control group has, in effect, been created and the techniques are again appropriate.

A second aspect of case-control studies, which expands the definition offered, is that the exposures of interest are not limited to environmental factors; the genotype and endogenous factors may be investigated suitably with the case-control design. Similarly, a case-control study may relate to factors other than possible etiological agents, including possibly protective factors. Studies of the relationship of oral contraceptives to benign breast diseases exemplify this (Vessey, Doll & Sutton, 1971; Kelsey, Lindfors & White, 1974). Indeed, it may prove possible to extend the case-control design far afield from etiological investigations to such subjects as the evaluation of a health service. For example, Clarke and Anderson (1979) recently attempted to evaluate the efficacy of the Papanicolou smear by the case-control technique.

Finally, it is worth mentioning that although many techniques of survey research (e.g., questionnaire construction, subject selection) are used in case-control and other epidemiological studies, these studies are not examples of survey research. No etiological investigation, whether epidemiological or experimental, need describe a population; and, in a case-control study, neither the cases nor the controls need be representative of any population as conventionally designated. It is useful to consider, however, that even a case-control study which is not population-based does derive from a hypothetical population, being those individuals who, if they were to develop the cancer under study, would be included as cases but are otherwise potential controls. It is important that the vast majority, and preferably all, of the cases genuinely have the specified disease and that the controls are comparable to them; comparability implies the absence of bias, especially selection bias and recall bias.

While the case-control design can be modified in many ways, discussion is facilitated if it is limited to an etiological investigation employing only two groups of subjects, and this monograph is so restricted.

History

In 1926 Lane-Claypon reported a case-control study of the role of reproductive experience in the etiology of breast cancer (Lane-Claypon, 1926). This appears to be the first case-control study of cancer (and possibly of any disease) which meets the definition offered above; in fact the study is remarkably similar to a modern investigation. Lane-Claypon does not describe how or why she came to adopt this approach. Thereafter, until about 1950, there were no further case-control studies – at least of a cancer – similar in quality to that of Lane-Claypon. The design came to be used, however, for the investigation of outbreaks of acute diseases. For example, a comparison would be made between individuals with a foodborne disease and well persons with respect to specific foods eaten at a common meal.

In 1950 two case-control studies which linked cigarette smoking with lung cancer were published (Levin, Goldstein & Gerhardt, 1950; Wynder & Graham, 1950); and in ensuing years there were numerous similar studies of many cancers. Of these the smoking and lung cancer investigation by Doll and Hill warrants mention as the prototype case-control study (Doll & Hill, 1952). The 1950s also brought the first studies of case-control methodology. Especially important was Cornfield's demonstration that the exposure frequencies of cases and controls are readily convertible into a parameter of greater interest to most public health workers, namely the ratio of the frequency of disease among exposed individuals relative to that among the non-exposed (Cornfield, 1951). This parameter has several different names and somewhat different interpretations depending on the particular type of cases used in a case-control study; however, it is now widely referred to as the relative risk and is so described in this monograph. Another major paper of the 1950s was the synthesis of Mantel and Haenszel, which clarified the objectives of case-control studies, systematized the issues requiring attention in their performance and described two widely-used analytical techniques, a summary chi-square statistic and a pooled estimator of the relative risk (Mantel & Haenszel, 1959). It is encouraging that in 1977 an enumeration of the citation of papers published in the Journal of the National Cancer Institute showed the Mantel-Haenszel paper in sixth place and increasing in use (Bailar & Anthony, 1977).

Current perceptions of the epidemiological aspects of case-control studies are presented in a recent paper by Miettinen (1976), in the related correspondence (Halperin, 1977; Miettinen, 1977) and in the proceedings of a symposium on the topic (Ibrahim, 1979). This monograph represents a synthesis of recent progress regarding statistical aspects.

Present significance

From the mid-1950s to the mid-1970s the number of case-control studies (not necessarily cancer-related) published in two general and two epidemiology-related medical journals increased four- to sevenfold. In the mid-1970s, they comprised 7% of all papers published (Cole, 1979). More specifically pertinent to cancer, the 1979 edition of the *Directory of On-going Research in Cancer Epidemiology* (Muir & Wagner, 1979) includes 1 092 research projects compared with 622 in the 1976 edition (Muir & Wagner, 1976). Of the 1 092 current projects 320 (29%) were classified as case-control studies,

while only 143 (13%) were classified as follow-up (cohort) studies. These figures make an impressive statement about the present and possible future role of the case-control study in cancer research. More persuasive, however, would be a favourable assessment of the results of case-control studies of cancer. While this has to be subjective and perhaps reflects only an individual point of view, it is contended that, with the exception of our knowledge of carcinogenic occupational exposures (attributable mainly to the perspicacity of clinicians and the results of non-concurrent follow-up studies), most of our epidemiological knowledge of cancer etiology was established or originated from case-control research. In the past few years alone, the case-control study has brought to light or improved our understanding of such associations as late first birth and breast cancer (MacMahon et al., 1970); diethylstilboestrol and vaginal cancer in young women (Herbst, Ulfelder & Poskanzer, 1971); exogenous oestrogens and cancer of the endometrium (Ziel & Finkle, 1975; Smith et al., 1975); alcohol and tobacco consumption and cancer of the oesophagus (Tuyns, Péquignot & Jensen, 1977); the hepatitis B virus carrier state and cancer of the liver (Prince et al., 1975; Trichopoulos et al., 1978); and the role of dietary factors in cancers of the stomach and large bowel (Haenszel et al., 1972, 1973, 1976; Modan et al., 1974, 1975).

Apart from their frequency and the importance of their results, there is a more direct justification for placing a high value on the role of the case-control study in cancer research: it will be indispensable for the foreseeable future. What could replace it? Experimental research can provide persuasive evidence of the carcinogenicity of some kinds of agents for animals, but a generalization is required before such evidence can be applied to man. For some agents, for example, 2-acetylaminofluorene, a potent bladder carcinogen for several animal species, the generalization is readily made by nearly everyone; for others, for example, saccharin, an animal carcinogen under special circumstances, the relevance to man is not clear. Furthermore, the experimental approach may prove to be nonpersuasive or even inapplicable to the study of the carcinogenicity of some aspects of human lifestyles. The concurrent follow-up (prospective cohort) study is too expensive and time-consuming to be done often or as an exploratory venture. The non-concurrent follow-up (retrospective cohort) study requires the good fortune of locating old information on exposure which is relevant to the question at hand. Furthermore, follow-up studies usually cannot address interaction and confounding because the necessary information does not exist or because too few subjects develop the cancer of interest.

It is not by chance that the case-control study developed rapidly in the 1950s and is so popular today. Rather, the case-control study is contemporaneous with, and results from, the emergence of the chronic diseases as major public health problems requiring etiological investigation. The case-control study is uniquely well-suited to the study of cancer and other diseases of long induction period, for it permits us to look back through time from effects to causes. This is the reverse of the observational sequence of experimental research and of follow-up studies whether concurrent or non-concurrent. Nonetheless, the case-control study needs no apology since it is not backward, unnatural or inherently flawed. Indeed, recent applications of case-control selection procedures and analytical methods to follow-up studies show that the same results are achieved as from the analysis of the whole cohort but that costs are reduced and efficiency improved (Liddell, McDonald & Thomas, 1977; Breslow & Patton, 1979). Furthermore, in every-

day human affairs cause-effect relationships are often viewed in reverse temporal sequence, but there is no difficulty in recognizing them. Everyday affairs, however, usually have causal pathways that are short, simple and strong. When a causal pathway spans decades our ordinary perceptions may not suffice, and this is particularly true if the pathway is rather faint because the cause-effect relationship is weak, which it often is for cancers. Thus, we had to develop a more refined method of observation to look back through time; that refined and still evolving method is the case-control study.

None of the foregoing support of the case-control study deprecates other forms of research, experimental or non-experimental. Nor is it contended that the case-control study is flawless; poor case-control studies have been and will continue to be done, and even a well-designed and conducted case-control study may produce an erroneous result. Considering the frequency with which case-control studies are done, and the ease with which such studies can be launched for exploratory purposes, it is to be expected that some contradictory results will appear. In this respect, the case-control study is no different from other forms of research, including rigorous experimentation. Thus it seems inappropriate to use a smattering of conflicting results from case-control studies to justify the position that "Certain scientific problems of case-control studies are inherent in (their) architecture..." (Horwitz & Feinstein, 1979), especially when the "problems" are not described. On the other hand, most of us recognize that the case-control design is young and underdeveloped and that it presents many problems and challenges (Feinstein, 1979). Most would also agree with the participants in the Bermuda Case-Control Symposium that research into the case-control method *per se* should be encouraged and that a set of standards for such studies should be developed (Ibrahim, 1979 [Discussion]). These constructive suggestions reflect the realization that the case-control study is different from, and more complex than, most experimental research designs and that some criteria for a good experiment are not only irrelevant to it but would be counter-productive. Criticisms of the case-control design (Sommers, 1978) which appear to reflect a judgement based on criteria for experiments should not be accorded.

1.2 Objectives

The principal objective of an etiological case-control study is to provide a valid, and reasonably precise, estimate of the strength of at least one hypothesized cause-effect relationship. In practice, this objective is usually supplemented by several others. The more common of these are the evaluation of several hypotheses and the description of the circumstances under which the strength of a cause-effect relationship varies, that is, of biological interaction. These objectives are identical to those of follow-up studies and even of experimental investigations of etiology.

The identity of objectives of case-control studies and of, say, experiments emphasizes two important things. The first is that generalizability of results is not a principal objective, while validity is. This is an important distinction to bear in mind since the two objectives validity and generalizability can be in competition. To illustrate this, the validity objective suggests that a case-control study should be based on a rather narrowly-defined case series and on controls highly comparable with them. For example, rather than including all women with breast cancer a study could be restricted to pre-

menopausal cases (and controls). Subject restriction of this type mimics the experimental situation in which homogenous animals are used in an effort to improve efficiency and to reduce the prospect that confounding could explain the results.

On the other hand, the wish to achieve the secondary objective of generalizability would result in an effort to identify all cases of a disease occurring in a designated population and to use a random (or stratified random) sample of that population as controls. Two considerations should be kept in mind before generalizability is sought. First, if the subjects are highly heterogeneous, the results for any subgroup are likely to be imprecise because of random variation. This imprecision leads to a lack of confidence in the validity of the results which, in turn, precludes generalization. Second, the willingness to generalize is ultimately subjective and dependent on knowledge of the subject matter, particularly of whether susceptibility to the cause is likely to differ between the group studied and the group to which one would like to generalize. Furthermore, few of us are willing to generalize the results of etiological research until there have been similar findings from at least two studies done in different demographic settings. These considerations suggest that validity should not be compromised in an effort to achieve generalizability; generalizability will follow from a valid result and especially from a series of valid results.

The second thing which follows from the similarity of objectives of case-control studies and experiments is the desirability of expressing results in terms which have a biological meaning and interpretation. In practice this means providing a measure of the difference, if any, in the frequency of disease between exposed and non-exposed persons, including, if possible, a description of the dose-response relationship. The measure to be provided is the relative risk and, if possible, the (absolute) difference in incidence rates or prevalences between exposed and non-exposed individuals. It is insufficient to provide only the exposure frequencies of the cases and controls with the related p-value or to provide only the coefficients and p-values derived from a multivariate model.

1.3 Strengths

The major strength of epidemiology compared with experimental research is that it applies directly to human beings. In an epidemiological study there is no species barrier to overcome in attempting to infer how applicable the results are to man. The major strength of the case-control design compared with other types of epidemiological research is its "informativeness". A case-control study can simultaneously evaluate many causal hypotheses whether they have been previously evaluated or are new. These studies also permit the evaluation of interaction – the extent and manner in which two (or more) causes of the disease modify the strength of one another. This design also permits the evaluation and control of confounding, that is, of an association resulting because the factor under study is associated with a known or suspected cause of the cancer. The reason for the informativeness is the large number of ill persons who are observed in a case-control study. In a follow-up study usually only a few subjects develop any one cancer. The others contribute relatively little information.

There is another way in which a case-control study is highly informative. If a population-based series of incident cases has been assembled, it is possible to describe the picture of the disease in that population. That is, one can describe incidence rates ac-

cording to age, sex and various risk factors at a point in (in fact during a brief period of) time. This cannot be achieved in a follow-up study, even if a general population comprises the study group (which is uncommon), unless new sub-cohorts are periodically added to the persons under observation. The reason is that a follow-up study group gives a broad picture of a cancer only at the start of the study. Thereafter, the group evolves and certain subgroups, e.g., the young, "disappear"; another subgroup literally disappears from the study – those lost to follow-up. On the other hand, the population-based case-control study provides a "window" on the totality of a cancer. One example of such a study relates to cancer of the bladder (Cole, 1973).

A second advantage of the case-control design is its efficiency, which is particularly impressive in view of its high informativeness. Such studies may be done in a few weeks if pre-existing data are used, but more often they take a year or two especially if the subjects are interviewed. Furthermore, the cost of these studies has tended to be low because only pre-existing or anamnestic data were gathered on relatively few subjects; case-control studies usually include several hundred subjects compared with the many thousands in follow-up studies. However, this low cost is becoming less characteristic of case-control studies of cancer because the kind of data required is changing. Many studies now require that interviews be supplemented with complex biochemical or other types of analysis of biological specimens. Despite the increased cost, this change is welcome for it is due to improvement in our understanding of the causes of cancer. On the other hand, the use of biological specimens in case-control studies may sometimes contribute nothing, or even be inappropriate, because the changes found may reflect some pathophysiological effect of the cancer rather than a cause. The advantages, then, of speed and low cost, while a general attribute of case-control studies, are not characteristic of all of them. Moreover, speed and low cost are not unique to the case-control study, and indeed the non-concurrent follow-up study is usually superior in these aspects.

A third advantage of the case-control study is its applicability to rare as well as common diseases. In this context, all but the three most common cancers (those of the breast, lung and colon in the western world) are "rare". In addition, the more rare the cancer, the greater is the relative advantage of this design. A disease which is rare in general, however, may not be rare in a special exposure group. If this is suspected and if such a group can be identified from a period in the past, a non-concurrent follow-up study should be considered.

A fourth advantage is that case-control studies (as well as follow-up studies) permit evaluation of the causal significance of a rare exposure. This is often not appreciated, and it is a common misconception that a case-control study is inappropriate for study of a rare exposure. Insofar as the prevalence of an exposure makes it suitable or unsuitable for a case-control study, it is not the general prevalence (i.e., among potential controls) but that among the cases that is relevant. If a factor is rare, but nonetheless accounts for a high proportion of the cancer, that is, if the factor is related to a high population-attributable risk percent (Cole & MacMahon, 1971), it can be studied. Indeed this is a circumstance that maximizes the efficiency of the case-control design since it enables a small study to be quite powerful. One example is the case-control study of eight young women with clear-cell adnocarcinoma of the vagina and 32 controls (Herbst, Ulfelder & Poskanzer, 1971). Similarly, a large fraction of mesotheliomas of the pleura are related to exposure to asbestos (Greenberg & Lloyd Davies, 1974), and

benign hepatomas in young women are related to use of the contraceptive pill (Edmondson, Henderson & Benton, 1976). On the other hand, a common exposure may prove unsuitable for a case-control investigation if it accounts for a small proportion of the cancer; in this situation a very large study will be required.

1.4 Limitations

One limitation suggested to characterize case-control studies is that they permit estimation only of relative disease frequency. This requires qualification. If a case-control study includes, as many have, an incidence or prevalence survey, it can provide risk factor-specific absolute measures of cancer frequency. For example, Salber, Trichopoulos and MacMahon (1969) provided incidence rates of breast cancer by marital status, and others have provided incidence rates of bladder cancer according to cigarette smoking status (Cole et al., 1971) and occupation (Cole, Hoover & Friedell, 1972). Even when a survey is not included it may be possible to estimate the absolute overall frequency of disease among the types of subjects studied and to infer risk factor-specific absolute frequencies. This is especially so with respect to cancer because of the information on incidence rates available from cancer registries. A method for doing this is described by MacMahon and Pugh (1970), and an example is the study of oral contraceptives and thromboembolic and gall-bladder disease from the Boston Collaborative Drug Surveillance Program (1973).

A second proposed limitation of case-control studies remains correct and is important. Namely, that these studies are highly susceptible to bias, especially selection bias which creates non-comparability between cases and controls. The problem of selection bias is the most serious potential problem in case-control studies and is discussed below. Other kinds of bias, especially that resulting from non-comparable information from cases and controls are also potentially serious; the most common of these is recall (anamnestic or rumination [Sackett, 1979]) bias which may result because cases tend to consider more carefully than do controls the questions they are asked or because the cases have been considering what might have caused their cancer. The weakness then of case-control studies is that, in the end, the investigator must appeal to subjective or only semi-quantitative arguments to the effect that the information that he has from cases and controls is equivalent in source and quality. Thus, to a great extent the problem of doing a persuasive case-control study is that of avoiding bias. In one sense this is a basis for optimism because the sources and nature of biases in epidemiological studies have only recently come under scrutiny (Sackett, 1979), and we can expect progress in developing methods for their identification and control. Yet, there will be biases peculiar to each cancer and to each exposure and even to each study. It may be possible, at least, to define the more important biases that commonly affect certain kinds of case-control studies; Jick and Vessey (1978) have attempted this for case-control studies of drug exposures.

1.5 Planning

The case series

When a problem has been defined and a case-control study decided upon, attention is usually given first to designating the cases. The goal should be to designate a group of individuals who have a malignancy which is, as far as possible, a homogeneous etiological entity. It will be easier to unravel a single causal web rather than several at one time. For example, only a very limited level of knowledge can be reached in the study of the epidemiology of "cancer of the uterus"; but, if adenocarcinoma of the uterine corpus is distinguished from squamous-cell carcinoma of the cervix and if research is directed at one or the other, considerable progress can be made. We can go further in making these distinctions by not defining diseases solely in terms of manifestational characteristics, no matter how refined we consider them to be. Definitions of disease based on their clinical or histological appearance suffice when there is a one-cause/one-manifestational-entity relationship. But they do not suffice for cancer. To study the complex cause-effect relationships of cancer we should attempt to use all existing knowledge, manifestational and epidemiological, to designate a restricted case series which is as homogeneous as possible with respect to etiology. The inadequacy of using only organ site and histological appearance or cell type to specify an etiological entity is made clear from one of the ideas about multiple-causation. In this, it is suggested that one type of cancer may have more than one independent set of causes. In order to identify cases likely to have the same set of causes the case series could be restricted according to age, sex, race or some other appropriate factor.

The restriction of case characteristics may bring benefits besides providing a series homogeneous with respect to cause. For one, the narrower range of possible causative factors is more likely to exclude false causes from consideration – "causes" which turn out to have no association with the cancer, or an association due only to confounding.

Another and especially important benefit is that the problem of selection bias may be reduced. When there is no association between the factor and the cancer of interest, there are nonetheless many ways in which an association may appear in the data of a case-control study. (The reverse situation, of no apparent association when in fact there is one, may also occur for similar reasons, but is not illustrated here.) One of the most important ways in which a false association can be created is by a selection bias. The question of selection bias must be considered simultaneously for both the case and the control series, since it is a question of their comparability; however, the problem of selection bias can most easily be appreciated with reference to case selection. Some mechanisms of selection bias may best be minimized by appropriate methods of case selection, thus the topic is presented here. By selection bias, I mean the bias which results when cases (or controls) are included in (or excluded from) a study because of some characteristic they exhibit which is related to exposure to the risk factor under evaluation. Often, for cases, the characteristic will be a sign or symptom of the cancer under study which is not always due to the cancer. This definition makes it clear that selection bias is not a single or simple phenomenon. It may, for example, represent a selection force towards inclusion in the study which operates on cases or on controls, or on both but unequally. Nor is this selection a conscious one; indeed, the parties

applying the force will usually not even be aware (because it is usually not known) that the selection characteristic is associated with the factor under study. An example of a proposed selection bias follows. In 1975 two groups reported a rather strong association between the use of exogenous oestrogens and cancer of the endometrium (Ziel & Finkle, 1975; Smith et al., 1975). Later, Horwitz and Feinstein (1978) proposed that the association was due to a selection bias. They pointed out that women who use exogenous oestrogens are more likely than those who do not to experience vaginal bleeding, a moderately frequent but not serious side effect of the medication. However, vaginal bleeding in a postmenopausal woman is a matter of concern since it is a common sign of cancer of the endometrium. Thus, a postmenopausal woman who exhibits vaginal bleeding is likely to be examined closely for endometrial cancer whether or not she takes oestrogens. Usually, this includes a histological examination of tissues taken from her endometrium. The basis for this proposed bias is complete if one accepts the suggestion that in a high proportion of apparently normal postmenopausal women there is a condition which, morphologically, mimics cancer of the endometrium or which "is" cancer of the endometrium of an indolent type. Thus, Horwitz and Feinstein (1978) proposed that the use of oestrogens was merely attracting attention to a large number of these indolent conditions by causing vaginal bleeding and diagnostic evaluations. (If correct, this would serve as an excellent example of selection bias. However, Hutchison (1979) reviewed the Horwitz-Feinstein proposal and concluded on several bases that, while the proposed selection bias is credible and may even occur, it is unlikely to be sufficiently strong to account for any but a small part of the approximately sevenfold excess risk of endometrial cancer among oestrogen users. The fact that the association persists when the base series is restricted to women with frankly invasive cancer (Gordon et al., 1977) supports this view).

This type of selection bias is not likely to be a problem in case-control studies of cancer. For one reason, virtually by definition, cancer tends to be a progressive condition which ultimately comes to diagnosis. The bias is much more likely to operate in studies of non life-threatening conditions which produce no, or tolerable, symptoms. The bias is also more likely to operate in studies of drug exposure than in those of etiological agents in general, for drugs produce many adverse effects some of which cause a patient to be subjected to a battery of diagnostic tests.

The status of cases to be included in a case-control study must also be decided. There are three types that are often used – incident, prevalent and, occasionally, decedent. The use of decedent cases will not be discussed except to point out that their study brings the same problems as the study of prevalent cases plus additional ones. Decedent cases probably should not be used except in a preliminary study of a disease based on medical record review or in the study of a disease which becomes manifest by causing sudden death.

Incident or newly-diagnosed cases are to be preferred and are the type usually used in case-control studies of cancer. They have several advantages over prevalent (previously diagnosed) cases. For one, the time of disease onset is closer to the time of etiological exposure than is any later time. Thus, an incident case should recall better than a prevalent case the experience or exposures under evaluation. Similarly, recent medical, employment or other records are likely to be available and more informative than older records. In addition, a series of incident cases has not been acted upon by the

determinants of survival whereas a series of prevalent cases has. That is, the cases prevalent at any point in time are the survivors of a larger series of incident cases diagnosed in a preceding period. If incident and prevalent cases differ with respect to risk factors, the use of prevalent cases would give an erroneous result. For example, using prevalent cases there appeared to be an association between the HLA antigen A2 and acute leukaemia (Rogentine et al., 1972). A second study, however, showed that this association was due to improved survival among cases with HL-A2, rather than to an increased risk of developing the disease among persons with HL-A2 (Rogentine et al., 1973). A third advantage of incident cases is that the effects of the disease are less likely to appear as causes. If a case has been prevalent for several years he may have changed his environment or lifestyle in a number of ways. Then, unless care is taken to restrict inquiry to the pre-morbid circumstances, a false characterization will occur. A final advantage of incident cases is that they relate more directly to the usual objective of an etiological investigation; i.e., with incident cases one evaluates the way in which exposure relates to the incidence rate of a cancer not to its prevalence.

There is only one advantage to the use of prevalent cases over the use of incident cases: they are already available. This might be considered a major advantage, since most cancers are sufficiently uncommon that it could take several years of ascertainment at several medical centres to assemble an adequate number of incident cases. However, the case-fatality of cancer remains sufficiently high that, usually, a large series of prevalent cases cannot be assembled unless cases diagnosed long ago are included. This may provide abundant opportunity for determinants of survival to act. For these reasons there is usually no appreciable advantage to the exclusive use of prevalent cases.

When the case series has been defined in terms of characteristics and type, a source must be located. In most case-control studies cases are identified by monitoring of pathology department log books, hospital operating-room schedules, or discharge lists. Less frequently, the office records of a number of physicians are used. In cancer research, an additional source of cases is the hospital or regional cancer registry. However, unless a special effort is made, regional cancer registries usually cannot identify cases until three months or more after diagnosis.

The control series

The designation of the type, number and size of the control group or groups and the problem of selecting the specific control subjects are perhaps the most important and most difficult tasks in planning a case-control study. The issue of the number of control groups was addressed above. Here the issues of the type and size of the group are discussed. The method of selecting the individual control subjects will not be discussed as it is almost entirely dependent on study-specific circumstances. One general question, however, relates to whether, when using a hospital-based control series, subjects with conditions known to be associated with the exposure under study should be eligible as controls. Most epidemiologists consider it reasonable to exclude them if the exposure-related illness is the reason for their current hospitalization.

There is no one type of control group suitable for all studies and, it must be acknowledged, there are no firm criteria for what is an acceptable group. The major factors which contribute to choice of a control group are the characteristics and source of the

cases and knowledge of the risk factors of the cancer to be studied and of how these might confound or interact with the exposure to be investigated.

The characteristics and source of the case series must heavily influence the type of control selected if comparability of the two series is to be achieved, that is, if selection bias is to be avoided. In general, if a population-based series of cases has been assembled, a random or (age and sex) stratified random sample of the population will prove to be a suitable control series. If cases have come from a restricted source it is usually appropriate to select controls from the same source. In an extension of the latter notion, Horwitz and Feinstein (1978) suggested that reduction of selection bias would be achieved if controls were selected from among persons who had undergone the same diagnostic test as the cases and found not to have cancer. This is intended to overcome bias due to selecting cases from among those who are excessive users of medical services. Such people may be highly likely to have had diagnostic tests performed, even on marginal indications, and may also have an unusual exposure experience. However, if controls are also drawn from those who have had the diagnostic test, then they should be more closely comparable with cases in terms of medical service use and exposure experience. An example of this approach to the control of selection bias is the study by Horwitz and Feinstein (1978) of cancer of the endometrium; controls were drawn from among women who had been evaluated by uterine dilatation and curettage, just as the cases had been. This type of control group would appear inappropriate because agents which cause one disease in an organ often cause other diseases of that organ, or signs or symptoms referable to it. If such a procedure is followed in a study of, say, lung cancer, individuals with chronic pulmonary diseases would comprise a high proportion of the control series. An association of lung cancer with cigarette smoking would still be perceived, because it is a strong association, but it would be muted because smoking causes many diseases of the lung. Despite this difficulty, the use of a diagnostic register as a source of controls may be a useful way to reduce the possible "medical consumerism" bias described above. However, to be appropriate, such rosters of potential controls should relate to procedures for the diagnosis of conditions of organs other than that organ which is the site of the cancer afflicting the cases.

It has been suggested, for yet another reason, that the control series should have undergone or be subjected to the same diagnostic procedures as the cases. The reason proposed is that it would permit exclusion of early cases or "cases-to-be" from the control series and thus permit a more appropriate comparison. This exclusion is contrary to principle since even cases-to-be are a portion of the at-risk population (whether a real or hypothetical population), and their exclusion would distort the estimated frequency of exposure among the group as a whole (Miettinen, 1976). The exclusion is also difficult to accept in practice since it would be expensive, would pose practical problems, and for some procedures would be ethically unacceptable. Furthermore, since the remaining lifetime risk of developing any specific cancer is 10% or less (for most cancers much less) very few potential controls would be excluded in this way.

Another consideration in designating the control series is related to the persistent opinion that the controls must be like the cases in every respect apart from having the disease of interest. The historical basis of this misconception is clear; it comes from the axiom of experimental research that control subjects must be treated in every respect

like exposed subjects. But in a case-control study this axiom is inapplicable. The consequences of selecting controls to be like cases with respect to some correlate of the exposure under study but to a correlate which is not a risk factor, that is, of "overmatching", are now recognized and are discussed in the section on matching.

A second aspect of control selection is the size of the series. When the number of available cases and controls is large and the cost of gathering information from a case and a control is about equal, then the selection ratio of controls to cases would be unity. The standard issues would then be invoked to estimate the acceptable minimum study size. The question becomes more complex when the size of either group is severely limited or the cost of obtaining information is greater for either cases or controls. For example, it occurs frequently that only a small number of cases is available for study. In this circumstance, the selection ratio should be increased to two, three or even four controls per case. This is not commonly done, and it is regrettable to see otherwise good case-control studies which are non-persuasive because of the unnecessarily small size of the control series. The selection ratio should be permitted to vary according to the circumstances of the study. But, one must be wary; it is wise to stay within the bounds of 4:1, except when the data are available at little cost or when they were collected for another purpose, and especially if they are in the form of a machine-readable file. The reasons for this have been presented by Gail et al. (1976) and by Walter (1977) and relate mainly to the small increase in statistical power as the ratio increases beyond four.

A third aspect of the designation of controls and a major factor in case-control studies is the source of the control group. Most studies use either hospital patients or the general population; restricted population groups, e.g., neighbours of cases or special groups such as associates or relatives of cases are much less often used.

The general population has a major strength as the source of the control series. Such controls will be especially comparable with the cases when a population-based series of cases has been assembled. This often makes for the most persuasive type of case-control study because of the high comparability of the two series and because a high level of generalizability of results is achieved. Even when a hospital-based series of cases is assembled, the population controls have the attribute of being, in general, well, and so causes of disease are not inordinately prevalent among them. Thus, one usually need exclude nobody from a population control series. (There are some exceptions to this. For one, it is reasonable to exclude people who do not have the organ in which the cancer develops. This is of some significance in studies of cancer of the uterus, at least in the United States where 30% or more of women aged about 50 years do not have a uterus. For another, it seems reasonable to exclude a control who has been previously diagnosed with the cancer under study.) There are, however, three disadvantages associated with use of the general population as a control group. Firstly, it can be extremely expensive and time-consuming to select and to obtain information from such a group. Secondly, the individuals selected are often not cooperative and response tends to be worse than that from other types of controls. This second disadvantage is especially important because it detracts from the presumed major strength of a general population control group. A third disadvantage of a population-based control series may arise if it is used in the study of a disease with mild symptoms for which medical attention need not be sought. In this instance the factors which lead to seeking medical care, such

as, perhaps, affluence, will appear to be correlated with the disease. This problem is a small one in studies of cancer in countries where medical care is generally available.

Other kinds of general population control groups besides a random sample are sometimes used. Probably the most common of these is a neighbourhood control series. If these controls are obtained by having the interviewer actually move physically through each neighbourhood, the cost of the study may be extremely high. Furthermore, it may be difficult or impossible to characterize or even to enumerate the non-respondents (not at home) or the non-cooperators (those who decline to participate). It appears that non-cooperation is high when this approach is used. A recent example of a study using a neighbourhood control group is that of Clarke and Anderson (1979). For each control finally obtained, an average of 12 household contact efforts were required, one of which led to a non-cooperator. Thus, the effective proportion of cooperators in this study was about 50% and even that was obtained from among a self-selected group of respondents. On the other hand, if neighbourhood controls can be selected by use of some type of directory or listing and the initial contact made by telephone or letter, response should be acceptable. Even so, it is usually difficult to accept the rationale offered to justify the use of close neighbourhood controls. Generally, this is stated to be the wish to match the controls to the cases with respect to socioeconomic class. But, people who live in the same neighbourhood are likely to be similar in more respects than socioeconomic class and so the potential for overmatching is great. A random sample of the general population is usually less costly to obtain and may be superior as well. If a factor such as socioeconomic class is to be controlled, this can be done in the analysis of a study using controls from the general population, provided the relevant information is obtained.

The use of hospital patients as a control group has several advantages. Such people are readily available, have time to spare and are cooperative. Moreover, since they are hospitalized (or have been recently) they may have a "mental set" similar to that of the cases. This should reduce anamnestic bias, one of the most serious potential problems in a case-control study. The use of hospital patients as controls may also make the cases and controls similar with respect to determinants of hospitalization. This is probably useful if the cases have a disease for which hospitalization is elective. Probably it is not very important in the study of cancer, unless the case series is assembled from one, or a few, highly specialized institutions which have a wide referral area. The use of hospital patients as controls has one possibly serious limitation. The controls may be in hospital for a condition which has etiological features in common with the disease under study. To minimize this problem, controls should be selected from patients with conditions in many diagnostic categories. Another limitation of hospital controls has arisen only over the past few years, particularly in the United States. Before approaching hospital patients it is usually necessary to have the approval of a responsible physician or surgeon; this is becoming difficult to obtain, presumably because of growing concern about the confidentiality of medical information.

Matching

In planning a case-control study it must be decided whether the controls are to be matched to the cases and, if so, with respect to what factors and how closely. This

warrants careful consideration because the decision will have implications for virtually every subsequent aspect of the study. Furthermore, an inappropriate decision will prove costly in time and money and may lead to an unsatisfactory study result. The issues underlying the question of matching received little attention until about ten years ago; but recently there have been several efforts at clarification, for example, those of Billewicz (1965), Miettinen (1976), McKinlay (1977), and Rubin (1979). These efforts have provided an appreciation of the complexity of what at first appears to be a straightforward approach to improving a study. This is an overview of some basic considerations relating to matching. In addition to being restricted to the case-control setting, this presentation is limited in that it deals primarily with factors which are dichotomous and only with matched pairs. These limitations do not distort the essential issues and permit them to be expressed more simply. In Chapters 5 and 6 ordinal scale exposures and multiple- and variable-ratio matching are considered.

In a case-control study the main purpose of matching is to permit use of efficient analytical methods to control confounding by the factors matched for. By confounding I mean the factor of interest is associated with the cancer under study, but this association is at least to some extent, and possibly entirely, non-causal. The association occurs because the factor of interest (the confounded factor) is associated with a true cause (the confounding factor) of the disease. Confounding can be illustrated by the concern expressed about the relationship between exogenous oestrogens and endometrial cancer. Steckel (1976) suggested that among women who will develop cancer of the endometrium at a later age, a fairly high proportion might have a rather difficult menopause. This is reasonable since cancer of the endometrium is probably caused by some hormonal difficulty, as are the signs and symptoms of the menopause. It is also reasonable to suggest that women who have a difficult menopause would be more likely than others to seek medical attention and to receive treatment with oestrogen, which is often prescribed for menopausal problems. If this were true, then oestrogens would (validly) appear to be associated with endometrial cancer in a case-control study (or in a follow-up study, for that matter). The association, however, would be non-causal, being confounded by a true cause of endometrial cancer, the hormonal aberration, which also "causes" women to receive oestrogens. (This particular proposed "constitutional confounding" was chosen as an example because it is quite illustrative, but it is almost certainly not correct since: (1) there is a dose-response relationship between oestrogen use and the relative risk of endometrial cancer; (2) the incidence rate of endometrial cancer has risen concurrently with the increase in oestrogen use; (3) the strength of the association between oestrogen use and endometrial cancer is similar in populations which are very dissimilar in their frequency of oestrogen use; and (4) cessation of oestrogen use is followed by a reduction in endometrial cancer incidence.)

A simpler, and correct, example of confounding is the association of cancer of the mouth with the occupation "bartender". Mouth cancer is caused by excessive alcohol and tobacco consumption, both of which are relatively common among bartenders. As can be seen from these two examples, for a factor to be confounding, it must be associated *both* with the cancer under study (as a cause of the cancer must be) and with the exposure of interest.

Confounding can be controlled in the analysis of a study or it may be eliminated by design in one of several ways. Of these ways, matching is by far the most often used,

probably because it appears to be a direct and intuitive approach. In addition, when there are only dichotomous (exposed, non-exposed) factors under evaluation, the matched-pairs (one control per case) design permits a straightforward estimation of the relative risk and its statistical significance. These features probably explain why pair matching was the first technique widely used to control confounding and remains popular today. But since there are now effective ways to control confounding in the analysis of the data the desirability of matching warrants reassessment.

Matching in a case-control study is an attempt to mimic blocking in an experiment, that is, randomizing animals within categories of a factor known (or suspected) to influence the outcome under evaluation. However, any analogy between blocking and matching is false in one crucial respect. In experimental work, no matter how extensive the blocking, the investigator manipulates exposure to the factor under study (usually by randomization). This guarantees that the blocking factors will not be correlated with the exposure of interest. However, exposure is not manipulated in a case-control study, and so a matching factor, unlike a blocking factor, will be associated with any exposure which differs between cases and controls. This must include the exposure under study if the matching is justified. For, if the matching is justified it will be with respect to a confounding factor, necessarily a correlate of the exposure under study. This means that any mode of analysis which fails to accommodate the fact that the matching process has forced the controls to be more like the cases than they otherwise would be, with respect to the exposure of interest, will lead to an estimate of the relative risk which is too close to unity. And, if the matching has been carried sufficiently far by matching on several variables, the cases and controls will be virtually identical with respect to the exposure to be studied. Effectively, time and money will have been spent in a counterproductive effort; the study will provide no information or, worse, an erroneous result. Thus it is necessary to avoid overmatching, that is, matching for a variable which is related to the exposure under study but which is not an independent risk factor and so cannot be a confounding factor.

It is useful to distinguish between two types of overmatching. One type occurs when the investigator matches for a factor which is part of the mechanism whereby the factor under study produces cancer. As an example, consider the prospect of matching controls to cases for the presence or absence of endometrial hyperplasia in a study of exogenous oestrogens and endometrial cancer. Hyperplasia is a condition which is caused by exogenous oestrogen and which may progress to cancer. The controls will thus be made very like the cases in exposure history, and the data, even when appropriately analysed, will lead to a relative risk biased towards unity. The second type of overmatching relates to matching for a variable which is a correlate of the factor under study, not an independent risk factor and not a part of the causal mechanism. In this instance an appropriate analysis will provide an inferentially valid estimate of the relative risk. However, the study will be inefficient, that is, imprecise, and there may be little confidence in the estimate obtained. The way this inefficiency comes about is described below.

Even matching which is indicated can be expensive and may prolong the data-gathering phase of a study. The number of matching "strata" is one of the determinants of cost and this increases sharply as the number of variables increases. For example, if a study involves matching for sex (two categories) and age (say, five categories) there will be ten strata. If matching for religion (say, three categories) is added there will be 30

strata, and formation of matched pairs will become difficult. The addition of one more matching variable, bringing the total number of strata to a minimum of 60, will make the search for a matched control tedious even for the more common types of subject. And for the less frequent types it may prove impossible to form matched pairs, with the consequent exclusion of some cases from the study. In addition to the number of strata, the specific variables chosen for matching will also influence the cost and time necessary to do the matching. Efforts have been made to match even for variables for which information must be obtained by interviewing the potential controls; if a control does not match the case for which be was being considered, the cost of contacting and interviewing him will usually be wasted. Generally, rather than expending resources in following an elaborate matching scheme, it will prove more efficient to gather data from a reasonably large number of potential controls and to evaluate and control confounding when the data are analysed. This approach can be especially efficient if the range of the subjects is restricted (perhaps one sex and a narrow age range) and if advance information is available as to whether individuals meet the restricted characteristics.

Matching can be envisioned as an effort to increase the contribution (or informativeness) of each subject to the study. Thus, while there may be relatively few subjects in a matched study, any matched pair which is discordant for exposure (one member exposed, the other not) makes a moderately large contribution to the evaluation of the relative risk and its statistical significance. However, each matched pair which is exposure-concordant makes no contribution at all. This illustrates the need to avoid matching for factors which are correlates of exposure but which do not confound the association of interest. The effect of such matching is to create an excessive number of uninformative exposure-concordant pairs.

A final cost of matching may have to be paid when the data are analysed. The matching process requires that the data first be analysed with the matching taken into account. If a stratified analysis is used (see Chapter 5), control for the confounding effect of factors other than those matched for will lead to the elimination of much of the data. However, if it is necessary to control such factors it may be possible to demonstrate that, as is often so, only age and sex are pertinent as matching factors and that the matching can be ignored as long as age and sex are carefully controlled in the analysis or the results are derived for specific age-sex groups. The analysis may then proceed and the effect of an unmatched confounding factor controlled for. If regression methods are used in the analysis (see Chapter 7) unmatched and matched factors can be controlled directly.

Some of the problems associated with matching are illustrated by an unusual case-control study done to evaluate the hypothesis that tonsillectomy is associated with increased risk of Hodgkin's disease (Johnson & Johnson, 1972). The study included 85 persons with Hodgkin's disease and, as their controls, 85 siblings, each sibling being matched to the respective case for sex and for age as well as, inherently, for sibship. The study was interpreted as showing no association between tonsillectomy and Hodgkin's disease. It seemed likely to others, however, that although the study consisted of a control series closely matched to the cases for likely strong correlates of tonsillectomy, especially sibship, the matching had been ignored in the analysis (the analysis had not been described). The data were then reanalysed (Cole et al., 1973) and a relative risk of 2.1 with $p = 0.07$ was found – a positive result consistent with an earlier report. This

illustrates the need to accommodate the matching with an appropriate form of analysis. In fact the 85 matched pairs were a subset of a larger series of 174 cases and their 472 siblings. The reduction to the 85 matched pairs, presumably to control potential confounding by sex and age, had caused 74% of the available data to be discarded. When all the data were analysed, thus ignoring age and sex, the relative risk was 2.0 ($p = 10^{-4}$). The near identity of the two relative risks is evidence that, in this series of subjects, there was no confounding by age and sex and that matching for those factors was irrelevant and wasteful.

The following should be considered when matching is contemplated for a case-control study. First, matching is only justified for factors which are known or suspected to confound the association of interest; that a factor may be related to the exposure of interest is not sufficient justification for matching. Second, matching may also be justified for factors which could interact with the exposure of interest in producing disease, since it provides more efficient estimates of relative risk within subgroups homogeneous with respect to suspected interacting factors. Third, it is usually possible to justify the costs in time and money of matching for age, sex and nominal scale variables with a large number of realizations (sibship, neighbourhood). However, such nominal scale variables should be matched for only if they meet one of the first two criteria, and this is uncommon. Fourth, when it is decided to match for a factor the matching should be as close as possible, with expense being the constraint to making an ever tighter match. For example, age is usually matched for arbitrarily in (plus or minus) five- or ten-year units. Frequently, it would cost very little more to match on year of birth or perhaps two-year age units. With respect to matching for age in particular it would be appropriate to modify the closeness of the match to the age of the subjects studied. For children and young adults a very close match is indicated because experiences change rapidly at these ages and because a discrepancy of a given magnitude, say one year, is a relatively greater proportion of the lifespan than it is in middle or old age. Since the principal objective of matching is to eliminate a potential confounder as such, the tighter match is to be desired since it minimizes the prospect that there would be "residual confounding" within the matching strata.

1.6 Implementation

Information gathering

The methods and problems of gathering information for a case-control study, as for other studies, greatly depend on the locale in which the study is done and the information sources used (interview, postal questionnaire, medical or other type of record review). Only a few general suggestions are offered. A case-control study usually begins with the investigator seeking cooperation from several hospitals or medical practitioners. This usually amounts to a request to identify cases, and perhaps controls, from available records. At least in the United States, this cooperation is becoming more difficult to obtain for several reasons, the major one being concern about litigation by a patient who believes that confidentiality has been breached. It is remarkable how deep and widespread this concern is, considering the rarity of the problem. For example, in the Department of Epidemiology at the Harvard School of Public Health during a period

in which at least 15 000 subjects identified from hospital records were requested to provide information, there was no such litigation nor serious threat of it. Despite this experience, which seems typical of epidemiological research, it is often not possible to persuade hospital administrators to cooperate. When non-cooperation is anticipated it is useful to make initial contact with a hospital through a staff physician who supports the research.

The next stage is to abstract the medical records of the cases, and of potential controls, at locations where cooperation was received. If possible the items on the record abstract form should follow the sequence of the medical record, but this may not be possible if several different hospitals are involved. If the medical record abstract involves information pertinent to the exposures under evaluation, as occurs when the role of drugs is in question, it is important to blind the abstractor to the case-control status of the record. It is also important to delete from the record or to mask any information relating to exposures sustained after the case's cancer was diagnosed, and during the equivalent time for the controls. These things will usually prove difficult and will involve at least two people in the record abstracting process. Nonetheless, both are usually justifiable.

It is best to conduct interviews concurrently for cases and controls. This should minimize the likelihood that learning by those gathering the data will influence the results. It would also minimize any effects of short-term changes, such as those of the seasons, or of some unexpected publicity about the cancer or the factors under study.

It is often recommended that, when they are involved, there should be as few interviewers as possible, preferably one. The rationale of this is that it will introduce uniformity into data collection. But, if the quality of the work is poor or if the interviewer is biased, a study would be ruined by having only one interviewer. It seems wiser, when practical, to have several well-trained interviewers. When more than one interviewer is used, it may be informative to analyse the principal study factors on an interviewer-specific basis. A positive result based on information from only a small proportion of the interviewers would be a cause for concern.

There are several suggestions concerning interviewers which are sound but difficult to meet. One of these is to keep the interviewers, and all study staff, unaware of the principal hypothesis(es) under evaluation. But even if the investigator attempts this, the interviewers usually become aware of what is important from the interview form itself or from sources external to the study. Another suggestion is that the interviewers be unaware of the status (case or control) of each subject they interview. One way of doing this is to have one person arrange the interview and another conduct it. This may prove effective for interviews conducted in hospital but is rarely even possible for those done at home. The reason is that the subjects, both cases and controls, usually want more information than was given them about the objectives of the study and the reason for their inclusion. In order to answer such questions in an honest, even if ambiguous, way the interviewer usually has to know the subject's status. The need for ambiguity often arises because many physicians still do not want their patient told of the diagnosis of cancer. Another reason is that the interviewer does not know, or should not inform the subjects, about the specific purpose of the study. A third suggestion concerning interviewers can and should be met: each of them should deal with the same ratio of controls to cases as exists in the study as a whole.

When a postal questionnaire is used to gather data an effort should be made to make the form as simple as possible in both appearance and use. This can be implemented by aligning the various insets so that there are as few different margins as possible. It is also useful to make the format of the response as uniform as possible, e.g., all boxes to be checked or alternatives to be circled, but not both. The instructions to the subject should be as brief and clear as possible. There is no disadvantage, however, in terms of response frequency, in making the questionnaire itself as long as required, within reason, nor does there seem to be any disadvantage (in terms of response frequency) in using franked as opposed to stamped mail, nor in using second or third class as opposed to first class service. In general, it appears that age and socioeconomic status of the subjects are determinants of response (younger and better-off subjects respond better), while features of the postage and questionnaire are relatively unimportant (Kaplan & Cole, 1970).

In both an interview form and a postal questionnaire, each item should deal with one question – compound questions should be avoided. It is usually advisable, especially in an interview, to permit unstructured responses; in such cases the space for the response should be followed by an indication of the units in which the response is to be expressed. Prescribed ranges to classify a response (e.g., <2, 2–4, 4–6, 7+ years) should be avoided; it is rarely justified to degrade information in this way at the time of collection. Note also that the responses in this example are ambiguous: the second and third are not mutually exclusive. All forms should be tested repeatedly to remove ambiguities and queries which elicit vague or ambiguous responses. Completed interviews should be reviewed by an experienced supervisor and the interviewers informed of the assessment of their work; part of this process should be undertaken by the investigator.

Information management

Information management commences well when good information-gathering forms and high-quality gathering and editing procedures are used. One aspect which relates particularly to information management is the use of "self-coding" forms. This term used to refer to several different formats, including one requiring the subject or interviewer to select a response and enter a corresponding code into a designated space. Generally, mixing information gathering with information management in this way is ill-advised; it is conducive to error and may reduce rapport in the interview setting. It is better to have all the coding done by two (or more) people (including, if convenient, the interviewers themselves) in a setting free of the stress of information gathering.

The information gathered must be translated into a series of numeric (or, rarely, alphabetic) codes. Generally, one item in the code will correspond with one item on the form. Each code item should consist of a series of mutually exclusive, collectively exhaustive categories. Virtually every item will require the categories "other" and "unknown", and rarely is it justified to combine these. No code item should be a "derived variable", i.e., a variable whose value is determined from two or more other variables. Computing the value of a derived variable is done more accurately, objectively, and at lower cost, by a computer. Just as there is no reason to degrade information at the gathering stage there is no need to degrade it at the coding stage. In fact, modern

analytical techniques argue for the use of highly refined data throughout the study, degradation to conventional ranges being reserved for the presentation of the results.

The early responses should be encoded by a highly experienced person and, if necessary, the code modified to permit designation of unanticipated responses or for other improvements. The need for this will be minimal if the forms have been well-designed and tested. The early work of the coders should be checked carefully to reveal any systematic problem resulting from a misunderstanding of how certain responses are to be encoded.

It is still common to have information encoded by one person and checked by another. The information is then key-punched and key-verified and a file created on a tape (usually for storage or transportation) or on a disc (for analysis). This procedure may serve for studies which gather an enormous amount of information and which can tolerate a modest amount of error, as is usually true of a follow-up study. However, most case-control studies gather a relatively small amount of information at relatively high cost. For these it is better to have each form coded by two people working independently. The code sheets prepared by one are then key-punched and a file created; no key-verification is done. The file is then printed out as a listing which is as easy to read as possible: triple spacing between lines, blank spaces between items. This printout is then checked against the code sheet of the second coder and errors are resolved. In this way all errors are caught by a single checking procedure, including coding errors which are often missed by the conventional procedure. Finally, a few of each of the computer-generated, derived variables are checked against the values generated by one person. While these procedures seem tedious they are not much more so than the usual ones, and they virtually guarantee that analysis can proceed in the knowledge that, as far as is humanly possible, the disc file is an accurate image of the information gathered.

1.7 Interpretation

The interpretation of a study involves evaluating the likelihood that the result reflects: one or more biases in design or conduct, the role of confounding, the role of chance or the role of causality. An approach to interpretation is presented here which is similar to that presented in Chapter 3 but is less concerned with quantification.

The most common basis for suggesting that a case-control study has produced an erroneous (biased) result (a suggestion which, of course, usually comes from a reviewer, not from the investigators) is that subject selection was inappropriate. This usually implies a selection bias but may refer to inclusion of non-cases in the case series (rarely a problem in the study of cancer) or "cases-to-be" in the control series. A second common basis for proposing a biased result is that there has been a systematic error in data collection such as that due to recall bias or to the interviewers knowing the case-control status of the subjects. A third basis for suggesting error is that there may have been an inordinate amount of random error in the data gathering. This suggestion is commonly offered for studies in which no apparent association emerges in relation to a factor acknowledged to be difficult to describe or quantify, such as diet. A fourth basis for suggesting error is that an inappropriate analysis has been done. Often the critic will suggest that the results are in error and imply that it is because of one or more of these reasons. Though it is not commonly done, it would be far more constructive if, in addi-

tion to invoking one problem or another, the critic would go further and, in discussion with the investigators, attempt to determine in which direction and to what extent the study result might be altered by correction of the proposed flaw. It is not rare for the critic of a positive study to imply that the correct result is the absence of association and to defend his proposal on the basis of a perceived bias which, if it truly existed and could be corrected, would probably cause the study result to be even more strongly positive.

A second interpretation to be considered is that the study result reflects confounding by some known (or suspected) cause of the cancer. This interpretation, it should be remembered, does not imply that the results are false. Rather, it implies that a valid but non-causal association exists between the cancer and the factor under study. Until recently, efforts were made to exclude confounding as an explanation of study results by showing that the proposed confounding factor was not associated to a statistically significant extent with the cancer under study. While it is understandable how this approach came to be used it is now unacceptable. The question of confounding is not dealt with in this way. Instead it is necessary to show to what extent the relative risk changes, or does not change, when the effect of the proposed confounding factor(s) is controlled. This change (if any) in the relative risk is an index of the degree of confounding. The relative risk estimate which persists after control of the confounding factor is the one which describes the specific association at issue.

When an unconfounded estimate of relative risk is available, interpretation turns to the possible role of chance. The issue, of course, is the possible role that chance effects in subject selection may have played in producing the unconfounded, not the crude, estimate of relative risk. This is addressed by estimating the significance level, or p-value, associated with the difference observed between cases and controls in their exposure histories. If this value is small, say, less than 0.05, it is usually concluded that the role of chance is unlikely to explain the observed departure of the relative risk from unity. There is nothing wrong with this, but it is a rather limited way to describe the role of chance. The confidence limits of the relative risk are more informative, especially in a study which shows no association. Use of the p-value and use of confidence limits are not mutually exclusive, but there are objections to the use of the p-value alone (Cole, 1979).

Finally, interpretation moves to the prospect that a valid causal association would explain the results. Occasionally, the causal inference is made as a "diagnosis of exclusion". That is, if the result is not perceived as biased and not due to chance or confounding, then it must be causal. But causality has at least three positive criteria and these should be reviewed, in addition to excluding alternative explanations. The *strength* of the association relates to causality. Relative risks of less than 2.0 may readily reflect some unperceived bias or confounding factor, those over 5.0 are unlikely to do so. The *consistency* of the association is germane to causality. Is the association seen in all subgroups where expected, and is there a dose-response relationship? Both these considerations relate to internal consistency. The extent to which the study is externally consistent, i.e., consistent with previous reports, can also be evaluated to support or refute a causal inference. That is, when a similar finding appears in different, especially very different, settings the notion of causality is favoured even if only because alternative explanations are less credible. A third criterion of causality is *biological credibil-*

ity; is it understood in biological terms how the exposure under study could produce the cancer of interest? However, while pertinent, the response to this question is not especially convincing one way or another; it has proven all too easy to propose credible biological mechanisms relating most exposures to most cancers; and, on the other hand, the failure to perceive such a mechanism may reflect only our ignorance of the state of nature.

For the sake of completeness another criterion of causality is mentioned: this relates to how the *frequency* of disease changes when the proposed cause is removed from (or added to) the environment. No doubt the response to manipulation of the exposure is the most cogent type of causal argument, but it does not concern the investigator dealing with the results of a particular case-control study.

Finally, there are two further considerations to bear in mind when interpreting a result. First, as an alternative to the four interpretations discussed, it could be decided that a study is "unevaluable". This decision is usually arrived at by exclusion, that is, there may be no basis for placing confidence in any of the other interpretations. The most frequent situation occurs when a study has no detectable flaw but its results are consistent with a chance effect. While the judgement of "unevaluable" may be tenable, it does not mean that the study is in error or has no value. Unless the study is so small as to be hopelessly imprecise, it can still make a contribution, in the context of other studies, to evaluating the hypothesis in question. Secondly, it is useful to keep in mind that the interpretation decided upon is not immutable. An investigator and the scientific community may favour one interpretation today and a different one later, in the light of knowledge acquired in the interim.

REFERENCES

Bailar, J.C. & Anthony, G.B. (1977) Most cited papers of the Journal of the National Cancer Institute 1962–75. *J. natl Cancer Inst., 59,* 709–714

Billewicz, W.Z. (1965) The efficiency of matched samples: an empirical investigation. *Biometrics, 21,*623–643

Boston Collaborative Drug Surveillance Program (1973) Oral contraceptives and venous thromboembolic disease, surgically confirmed gall-bladder disease, and breast tumours. *Lancet, i,* 1399–1404

Breslow, N.E. & Patton, J. (1979) *Case-control analysis of cohort studies.* In: Breslow, N. & Whittemore, A., eds, *Energy and Health,* Philadelphia, Society for Industrial and Applied Mathematics, pp. 226–242

Clarke, E.A. & Anderson, T.W. (1979) Does screening by "Pap" smears help prevent cervical cancer? A case-control study. *Lancet, ii,* 1–4

Cole, P. (1973) *A population-based study of bladder cancer.* In: Doll, R. & Vodopija, I., eds, *Host Environment Interactions in the Etiology of Cancer in Man,* Lyon, International Agency for Research on Cancer *(IARC Scientific Publications No. 7),* pp. 83–87

Cole, P. (1979) The evolving case-control study. *J. chron. Dis., 32,* 15–27

Cole, P. & MacMahon, B. (1971) Attributable risk percent in case-control studies. *Br. J. prev. soc. Med., 25,* 242–244

Cole, P., Elwood, J.M. & Kaplan, S.D. (1978) Incidence rates and risk factors of benign breast neoplasms. *Am. J. Epidemiol., 108,* 112–120

Cole, P., Hoover, R. & Friedell, G.H. (1972) Occupation and cancer of the lower urinary tract. *Cancer, 29,* 1250–1260

Cole, P., Mack, T., Rothman, K., Henderson, B. & Newell, G. (1973) Tonsillectomy and Hodgkin's disease (Letter to the Editor). *New Engl. J. Med., 288,* 634

Cole, P., Monson, R.R., Haning, H. & Friedell, G.H. (1971) Smoking and cancer of the lower urinary tract. *New Engl. J. Med., 284,* 129–134

Cornfield, J. (1951) A method of estimating comparative rates from clinical data. Application to cancer of the lung, breast and cervix. *J. natl Cancer Inst., 11,* 1269–1275

Doll, R. & Hill, A.B. (1952) A study of the aetiology of carcinoma of the lung. *Br. med. J., ii,* 1271–1286

Edmondson, H.A., Henderson, B. & Benton, B. (1976) Liver-cell adenomas associated with use of oral contraceptives. *New Engl. J. Med., 294,* 470–472

Elwood, J.M., Cole, P., Rothman, K.J. & Kaplan, S.D. (1977) Epidemiology of endometrial cancer. *J. natl Cancer Inst., 59,* 1055–1060

Feinstein, A.R. (1979) Methodologic problems and standards in case-control research. *J. chron. Dis., 32,* 35–41

Gail, M., Williams, R., Byar, D.P. & Brown, C. (1976) How many controls? *J. chron. Dis., 29,* 723–731

Gordon, J., Reagan, J.W., Finkle, W.D. & Ziel, H.K. (1977) Estrogen and endometrial carcinoma. An independent pathology review supporting original risk estimate. *New Engl. J. Med., 297,* 570–571

Greenberg, M. & Lloyd Davies, T.A. (1974) Mesothelioma Register 1967–68. *Br. J. ind. Med., 31,* 91–104

Gutensohn, N. & Cole, P. (1977) The epidemiology of Hodgkin's disease among the young. *Int. J. Cancer, 19,* 595–604

Gutensohn, N., Li, F., Johnson, R. & Cole, P. (1975) Hodgkin's disease, tonsillectomy and family size. *New Engl. J. Med., 292,* 22–25

Haenszel, W., Kurihara, M., Segi, M. & Lee, R.K.C. (1972) Stomach cancer among Japanese in Hawaii. *J. natl Cancer Inst., 37,* 969–988

Haenszel, W., Berg, J.W., Segi, M., Kurihara, M. & Locke, F.B. (1973) Large-bowel cancer in Hawaiian Japanese. *J. natl Cancer Inst., 51,* 1765–1779

Haenszel, W., Correa, P., Cuello, C., Guzman, N., Burbano, L.C., Lores, H. & Muñoz, J. (1976) Gastric cancer in Colombia. II. Case-control epidemiologic study of precursor lesions. *J. natl Cancer Inst., 57,* 1021–1026

Halperin, M. (1977) Estimability and estimation in case-referent studies (Letter to the Editor). *Am. J. Epidemiol., 105,* 496–498

Herbst, A.I., Ulfelder, H. & Poskanzer, D.C. (1971) Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *New Engl. J. Med., 284,* 878–881

Horwitz, R.I. & Feinstein, A.R. (1978) Alternative analytic methods for case-control studies of estrogens and endometrial cancer. *New Engl. J. Med., 299,* 1089–1094

Horwitz, R.I. & Feinstein, A.R. (1979) Methodologic standards and contradictory results in case-control research. *Am. J. Med., 66,* 556–564

Hutchison, G.B. (1979) Re: "Estrogens and endometrial carcinoma" (Letter to the Editor). *New Engl. J. Med., 300,* 497

Ibrahim, M.A., ed. (1979) The case-control study: consensus and controversy. *J. chron. Dis., 32,* 1–144

Ibrahim, M.A. & Spitzer, W. (1979) The case-control study: the problem and the prospect. *J. chron. Dis., 32,* 130–144

Jick, H. & Vessey, M.P. (1978) Case-control studies in the evaluation of drug-induced illness. *Am. J. Epidemiol., 107,* 1–7

Johnson, S.K. & Johnson, R.E. (1972) Tonsillectomy history in Hodgkin's disease. *New Engl. J. Med., 287,* 1122–1125

Kaplan, S. & Cole, P. (1970) Factors affecting response to postal questionnaires. *Br. J. prev. soc. Med., 24,* 245–247

Kelsey, J.L., Lindfors, K.K. & White, C. (1974) A case-control study of epidemiology of benign breast diseases with reference to oral contraceptive use. *Int. J. Epidemiol. 3,* 333

Lane-Claypon, J.E. (1926) *A Further Report on Cancer of the Breast. Reports on Public Health and Medical Subjects 32,* London, Her Majesty's Stationery Office

Levin, M.L., Goldstein, H. & Gerhardt, P.R. (1950) Cancer and tobacco smoking. A preliminary report. *J. Am. med. Assoc., 143,* 336–338

Liddell, F.D.K., McDonald, J.C. & Thomas, D.C. (1977) Methods of cohort analysis: appraisal by application to asbestos mining. *J. R. stat. Soc. Ser. A, 140,* 469–491

McKinlay, S.M. (1977) Pair matching – a reappraisal of a popular technique. *Biometrics, 33,* 725–735

MacMahon, B. & Pugh, B. (1970) *Epidemiologic Methods,* Boston, Little Brown

MacMahon, B., Cole, P., Lin, T.M., Lowe, C.R., Mirra, A.P., Ravnihar, B., Salber, E.J., Valaoras, V.G. & Yuasa, S. (1970) Age at first birth and breast cancer risk. *Bull. World Health Org., 43,* 209–221

Mantel, N. & Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. natl Cancer Inst., 22,* 719–748

Miettinen, O.S. (1976) Estimability and estimation in case-referrant studies. *Am. J. Epidemiol., 103,* 226–235

Miettinen, O.S. (1977) Estimability and estimation (Letter to the Editor). *Am. J. Epidemiol., 105,* 498–502

Modan, B., Lubin, F., Barell, V., Greenberg, R.A., Modan, M. & Graham, S. (1974) The role of starches in the etiology of gastric cancer. *Cancer, 34,* 2087–2092

Modan, B., Barell, V., Lubin, F., Modan, M., Greenberg, R.A. & Graham, S. (1975) Low-fiber intake as an etiologic factor in cancer of the colon. *J. natl Cancer Inst., 55,* 15–18

Muir, C.S. & Wagner, G., eds (1976) *Directory of On-going Research in Cancer Epidemiology 1976,* Lyon, International Agency for Research on Cancer

Muir, C.S. & Wagner, G., eds (1979) *Directory of On-going Research in Cancer Epidemiology 1979,* Lyon, International Agency for Research on Cancer *(IARC Scientific Publications No. 27)*

Prince, A.M., Szmuness, W., Michon, J., Demaille, J., Diebolt, G., Linhard, J., Quenum, C. & Sankale, M. (1975) A case-control study of the association between primary liver cancer and hepatitis B infection in Senegal. *Int. J. Cancer, 16,* 376–383

Rogentine, G.N., Yankee, R.A., Gart, J.J., Nam, J. & Trapani, R.J. (1972) HLA antigens and disease: acute lymphocytic leukemia, *J. clin. Invest., 51,* 2420–2428

Rogentine, G.N., Trapani, R.J., Yankee, R.A. & Henderson, E.S. (1973) HLA antigens and acute lymphocytic leukemia: the nature of the association. *Tissue Antigens, 3,* 470–476

Rubin, D.R. (1979) Using multivariate matched sampling and regression analysis to control bias in observational studies. *J. Am. stat. Assoc., 74,* 318–328

Sackett, D.L. (1979) Bias in analytic research. *J. chron. Dis., 32,* 51–63

Salber, E.J., Trichopoulos, D. & MacMahon, B. (1969) Lactation and reproductive histories of breast cancer patients in Boston. *J. natl Cancer Inst., 43,* 1013–1024

Smith, D.C., Prentice, R., Thompson, D.J. & Herrmann, W.L. (1975) Association of exogenous estrogens and endometrial cancer. *New Engl. J. Med., 293,* 1164–1167

Sommers, S.C. (1978) *Postmenopausal estrogens and endometrial cancer: a pathologist's overview.* In: Silverberg, S.G. & Major, F.J., eds, *Estrogens and Cancer,* New York, Wiley, pp. 45–53

Steckel, R.J. (1976) Re: "Estrogens and endometrial carcinoma" (Letter to the Editor). *New Engl. J. Med., 294,* 848

Trichopoulos, D., Tabor, E., Gerety, R.J., Xirouchaki, E., Sparros, L., Muñoz, N. & Linsell, C.A. (1978) Hepatitis B and primary hepatocellular carcinoma in a European population. *Lancet, ii,* 1217–1219

Tuyns, A.J., Péquignot, G. & Jensen, O.M. (1977) Le cancer de l'oesophage en Ille-et Vilaine en fonction des niveaux de consommation d'alcool et de tabac. Des risques qui se multiplient. *Bull. Cancer, 64,* 45–60

Vessey, M.P., Doll, R. & Sutton, P.M. (1971) Investigation of the possible relationship between oral contraceptives and benign and malignant breast disease. *Cancer (Philad.), 28,* 1395

Walter, S.D. (1977) Determination of significant relative risks and optimal sampling procedures in prospective and retrospective comparative studies of various sizes. *Am. J. Epidemiol., 105,* 387–397

Wynder, E.L. & Graham, E.A. (1950) Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma. A study of six hundred and eighty-four proved cases. *J. Am. med. Assoc., 143,* 329–336

Ziel, H.K. & Finkle, W.D. (1975) Increased risk of endometrial cancer among users of conjugated estrogens. *New Engl. J. Med., 293,* 1167–1170